

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:07 ; Search time 299.73 Seconds

(Without alignments)  
260,436 Million cell updates/sec

Title: US-09-331-631A-3

Perfect score: 3532  
Sequence: 1 MAINTSNLCSLFLSLFLSL.....SSRSTKQOQPLVSLIDFVGF 666

Scoring table: BIOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPRREMBL\_15:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3532	100.0	666	10 Q9SP14	Q9SP14 macadamia i
2	3412	96.6	666	10 Q9SP15	Q9SP15 macadamia i
3	3215	91.0	625	10 Q9SP13	Q9SP13 macadamia i
4	1324.5	37.5	593	10 Q9SEW4	Q9SEW4 juglans reg
5	1191.5	33.7	810	10 Q9ZWI3	Q9ZWI3 cucurbita m
6	1036	29.3	525	10 Q43358	Q43358 theobroma c
7	895	25.3	582	10 Q03865	Q03865 zea mays (m
8	875	24.8	637	10 Q03678	Q03678 hordeum vul
9	851	24.1	613	10 Q9M3X6	Q9M3X6 pisum sativ
10	833	23.6	544	10 Q22120	Q22120 glycine max
11	818	23.2	545	10 Q41674	Q41674 vicia narbo
12	798.5	22.6	489	10 Q9SP11	Q9SP11 glycine max
13	770.5	21.6	488	10 Q49927	Q49927 pisum sativ
14	765	21.4	448	10 Q40873	Q40873 picea glauc
15	757	21.4	417	10 Q22121	Q22121 glycine max
16	754.5	21.4	463	10 Q41677	Q41677 vicia narbo
17	754	21.3	450	10 Q40844	Q40844 picea glauc
18	749.5	21.3	438	10 Q43626	Q43626 pisum sativ
19	746	21.1	486	10 Q9L0U7	Q9L0U7 arabidopsis

20	706	20.0	518	10 Q9M3X8	Q9M3X8 lens culina
21	635	18.0	461	10 Q9ZRG9	Q9ZRG9 oryza sativ
22	565	16.0	430	10 Q41115	Q41115 phaseolus v
23	561	15.9	430	10 Q43633	Q43633 phaseolus v
24	554	15.7	421	10 Q43632	Q43632 phaseolus v
25	552.5	15.6	414	10 Q41727	Q41727 zania furfu
26	541.5	15.3	232	10 Q9SQ49	Q9SQ49 rullingia ma
27	534.5	15.1	540	10 Q03866	Q03866 zea mays (m
28	527	14.9	239	10 Q9SQ07	Q9SQ07 abroma augu
29	527	14.9	423	10 Q43617	Q43617 phaseolus l
30	523.5	14.8	239	10 Q40913	Q40913 phaseolus l
31	522	14.8	239	10 Q9SQ42	Q9SQ42 theobroma b
32	521.5	14.8	236	10 Q9SQ41	Q9SQ41 theobroma c
33	521	14.8	238	10 Q9SQ32	Q9SQ32 theobroma s
34	520.5	14.7	236	10 Q9SQ07	Q9SQ07 theobroma s
35	520	14.7	239	10 Q9SQ04	Q9SQ04 theobroma v
36	520	14.7	239	10 Q9SQ35	Q9SQ35 theobroma c
37	520	14.7	239	10 Q9SV79	Q9SV79 theobroma c
38	518.5	14.7	236	10 Q9SQ40	Q9SQ40 theobroma c
39	517.5	14.7	236	10 Q9SQ36	Q9SQ36 theobroma g
40	517.5	14.7	236	10 Q9SQ34	Q9SQ34 theobroma m
41	516.5	14.6	236	10 Q9SQ37	Q9SQ37 theobroma m
42	515.5	14.6	236	10 Q9SQ48	Q9SQ48 theobroma m
43	515	14.6	239	10 Q9SQ43	Q9SQ43 theobroma b
44	510.5	14.5	236	10 Q9SQ33	Q9SQ33 theobroma m
45	508	14.4	235	10 Q9SQ38	Q9SQ38 theobroma g

#### ALIGNMENTS

```
RESULT 1
Q9SP14 PRELIMINARY: PRT: 666 AA.
AC Q9SP14;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OC NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUIT KERNEL.
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antiticrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia.";
RL Plant J. 0:0-0(1999).
DR EMBL: AF161884; Amd54245.1; -.
DR HSP: P02853; 2PH.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 1.
DR SEQUENCE 666 AA; 78243 MW; 0ECA22F8710F8A7B CRC64;
```

Query Match 100.0%; Score 3532; DB 10; Length 666;  
Best Local Similarity 100.0%; Pred. No. 2e-248;  
Matches 666; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAINTSNLCSLFLSLFLSLTVSLAESFPDQRFYECKRQCMOLETSQMRVCVSOCD 60
DB 1 MAINTSNLCSLFLSLFLSLTVSLAESFPDQRFYECKRQCMOLETSQMRVCVSOCD 60
QY KREEDIDMSKYDNDPQTDCCQRCRQDSGPRQDQYCORCKEICEEEYVNROR 120
DB KREEDIDMSKYDNDPQTDCCQRCRQDSGPRQDQYCORCKEICEEEYVNROR 120
QY DPOQYEQCOERQRRHETEPHMQTCQRCERYEKKRQKRYEEOREDEKYEEM 180
DB DPOQYEQCOERQRRHETEPHMQTCQRCERYEKKRQKRYEEOREDEKYEEM 180
```

QY	181	KEEDKKRPQOREVEDCRRREODEPRQOYOCORCKREDOORONRGSGDLINPQRSGSRY	240
Db	181	KEEDKKRPQOREVEDCRRREODEPRQOYOCORCKREDOORONRGSGDLINPQRSGSRY	240
QY	241	EEGEEKOSDNPYFDERSLSTRFTEEGHLSVLBNFYGRSKLRLAKNRYLVLLEAPNA	3000
Db	241	EEGEEKOSDNPYFDERSLSTRFTEEGHLSVLBNFYGRSKLRLAKNRYLVLLEAPNA	3000
QY	301	FVLPHLDADAILLVVGGRGALKMIHRDNRESYNLECGDVIKIPDAGTTPYLINRDNREL	3600
Db	301	FVLPHLDADAILLVVGGRGALKMIHRDNRESYNLECGDVIKIPDAGTTPYLINRDNREL	3600
QY	361	HIAKLLOTISIPGOVKEEFPAGGONPEPYLSTFESKELTEALNMQARLGLVIGQOREGV	4200
Db	361	HIAKLLOTISIPGOVKEEFPAGGONPEPYLSTFESKELTEALNMQARLGLVIGQOREGV	4200
QY	421	IISASQEIRELTRDSESRRMHILRRGESSRGPNLNFNRPLYSNKYGAAYEVKPEDYR	4800
Db	421	IISASQEIRELTRDSESRRMHILRRGESSRGPNLNFNRPLYSNKYGAAYEVKPEDYR	4800
QY	481	QLOQMDVSVFLANTTQSGMMGPFFNTSTKKVYVASEADVEMACPHLSRRHGRGRGKR	5400
Db	481	QLOQMDVSVFLANTTQSGMMGPFFNTSTKKVYVASEADVEMACPHLSRRHGRGRGKR	5400
QY	541	HEEEDVHYEDVKARLSKREAIIVPVGHPVYFVSSGNENLLFEAFGINAONHNEFLAGR	6000
Db	541	HEEEDVHYEDVKARLSKREAIIVPVGHPVYFVSSGNENLLFEAFGINAONHNEFLAGR	6000
QY	601	ERNVLOQIEPQAMELAFAPRKVEEELFNSODESIFPPGPRHOHQSSNSTKQOQPLVSI	6600
Db	601	ERNVLOQIEPQAMELAFAPRKVEEELFNSODESIFPPGPRHOHQSSNSTKQOQPLVSI	6600
QY	661	LDPVGF 666	
Db	661	LDPVGF 666	
RESULT 2			
Q9SPLS PRELIMINARY; PRN; 666 AA.			
AC	Q9SPLS		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	VICILIN PRECURSOR.		
GN	AMP2.		
OS	Macadamia integrifolia (Macadamia nut).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.		
OX	NCBI_taxid=60698;		
RN	111		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=NUIT KERNEL;		
RA	Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;		
RT	"A family of antimicrobial peptides is produced by processing of a 75S		
RL	globulin protein in Macadamia integrifolia kernels.";		
DR	EMBL: AF161883; AD54244.1; -		
DR	HSSP: P02853; 2PHL		
DR	INTERPRO: IPR001113; -		
DR	PFAM: PF00346; Seedstore_7s; 1.		
DR	SEQUENCE 666 AA; 78217 MW; C752B884B2DF0224 CRC64;		

	Query Match	96.6%	Score 3412;	DB 10;	Length 666;
	Best Local Similarity	96.1%;	Fred. No. 1e-239;		
	Matches 640;	Conservative 12;	Mismatches 14;	Indels 0;	Gaps 0;
QY	1	MAINTSNLCSLFLFLSLFLSLTTVSLAESEPFROEVEECKRCOMLETSGMKRCVSOCD	60		
bb	1	MAINTSNLCSLFLFLSLFLSLTTVSLAESEPFROEVEECKRCOMLETSGMKRCVSOCD	60		

QY	61	KREEDIDMSXYDNDODDQOTQOCQORRCROESGPPROOYCOBRCKEICEEERENROR	120
Db	61	KREEDIDMSKTDNODDQOTQOCQORRCROESGPPROOYCOBRCKEICEEERENROR	120
QY	121	DPQOQYEOCQKHCQNHETEPRHMTQCOQRERARYEKKRKQOKRYEEOQREDEBEKYEEM	180
Db	121	DPQOQYEOCQKHCQNRHETEPRHMTQCOQRERARYEKKRKQOKRYEEOQREDEBEKYEEM	180
QY	181	KREDDKRPQOREYEDCRRCEQOEPQOYQOCQKRCQOORQNRGDDLINPQRGSGRY	240
Db	181	KREDDKRPQOREYEDCRRCEQOEPQOYQOCQKRCQOORQNRGDDLINPQRGSGRY	240
QY	241	EEGEEKOSDNPYEPDERSLSTRFTREEGHISYLENPFYGRSKTLFALKNYRLYLEANPNA	300
Db	241	EEGEEKOSDNPYEPDERSLSTRFTREEGHISYLENPFYGRSKTLFALKNYRLYLEANPNA	300
QY	301	FVLPPLHLDADAALLVTGGRGALKMTHRDNRESYVLECGDVIIRIPAGTFFYLLINRDNNE	360
Db	301	FVLPPLHLDADAALLVTGGRGALKMTHRDNRESYVLECGDVIIRIPAGTFFYLLINRDNNE	360
QY	361	HAFLKLOTISPPGOVKKEFPAGGONPEPYLSTFESKELLEALNQAERLGVIGQOREGV	420
Db	361	HAFLKLOTISPPGOVKKEFPAGGONPEPYLSTFESKELLEALNQAERLGVIGQOREGV	420
QY	421	IISASQEOIRELTRDDESRRKMHIRRGESSRGYPNLFNKRPLYSNKYGOAYEVKPEDYR	480
Db	421	IIRASQEOIRELTRDDESRRKMHIRRGESSRGYPNLFNKRPLYSNKYGOAYEVKPEDYR	480
QY	481	QLOQMDVSVFTANITQSGMMPFNTSTKVVVYVASEADYEMACPHLSRHRGCRGGR	540
Db	481	QLOQMDVSVFTANITQSGMMPFNTSTKVVVYVASEADYEMACPHLSRHRGCRGGR	540
QY	541	HEEEDVHYEQVKARLSKREAIIVPVGHVPYVSSGMDNLLLFAGGINAONNHENFLAGR	600
Db	541	HEEEDVHYEQVKARLSKREAIIVPVGHVPYVSSGMDNLLLFAGGINAONNHENFLAGR	600
QY	601	ERNVLQOIEPQAMELAFAPKKEVEBELFNODESIFFPGRPROHOOSNSTKQOOLVSI	660
Db	601	ERNVLQOIEPQAMELAFAPKKEVEBELFNODESIFFPGRPROHOOSNSTKQOOLVSI	660
QY	661	LDVFGE 666	
Db	661	LDVFGE 666	
RESULT 3			
Q9SPL3 PRELIMINARY: PRT: 625 AA.			
AC	Q9SPL3	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 13, Last annotation update)		
DE	VICLIN PRECURSOR (FRAGMENT).		
GN	AMP2.		
OS	Macadamia integrifolia (Macadamia nut).		
OC	Eukaryota, Viridiplantae, Embryophyta: Tracheophyta, Spermatophyta.		
OC	Magnoliophyta, eudicotyledons: Proteaceae: Macadamia.		
OX	NCHI_TaxID=60698;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=NUC KERNEL;		
RA	Marcus J.P., Goulter K.C., Green J.L., Manners J.M.:		
RT	"A family of antimicrobial peptides is produced by processing of a 75S		
RT	globulin protein in Macadamia integrifolia."		
RL	Plant J. 0:0-0(1999).		
DR	EMBL: AF161885; AAD54246.1; -		
DR	HSSP: P02853; 2PHL.		
DR	INTERPRO: IPR001113; -		
DR	PFAM: PF00546; Seedstore_7s; 1.		
FT	NON_TER 1		
SEQ	SEQUENCE 625 AA; 73586 MW; 415808A89D370296 CRC64;		

RD	EMBL	U_010-0(15597);	
DR	EMBL	AF161885; AAD54246.1;	-
DR	HSSP	P02853; 2PHL	
DR	INTERPRO	IPRO0113;	-
DR	PFAM	PF00546; Seedstore_75;	1.
PT	NON_TER	1	
SQ	SEQUENCE	625 AA; 73566 MW; 415808A89D370296 CRC64;	

Query Match	91.0%:	Score 3215:	DB 101:	Length 625:
Best Local Similarity	96.6%:	Pred. No.1,9e-225:		
Matches 604:	Conservative 7:	Mismatches 14:	Indels 0:	Gaps 0:
OY	42	OCMOLETSQOMRCVCSOCDFRFEEDIDMSKYNODDPQTDCOCCORRCRQOESGPRQOY	101	
Db	1	OCMOLETSQOMRCVCSOCDFRFEEDIDMSKYNODDPQTDCOCCORRCRQOESGPRQOY	60	
OY	102	CORCKEICEEEEFENRQDPQOYECOCRCORHETEFRRHQTQOOCERRYEKKRMO	161	
Db	61	CORCKEICEEEEFENRQDPQOYECOCRCORHETEFRRHQTQOOCERRYEKKRMO	120	
OY	162	OKRYEEOORDEDEKYEERKKEEDNKRPORCYEDCRRCCEOEPRQOYQOCRCOEOR	221	
Db	121	OKRYEEOORDEDEKYEERKKEEDNKRPORCYEDCRRCCEOEPRQOYQOCRCOEOR	180	
OY	222	OHRCGDILINPORGSGRYEEDGEEKOSDNPVYFDEKSLSTRFTEBGTIVLENFYGRSK	281	
Db	181	OHRCGDILINPORGSGRYEEDGEEKOSDNPVYFDEKSLSTRFTEBGTIVLENFYGRSK	240	
OY	282	LRALKKNRVLVLEPNPAFVLPETHADAILLVIGGRALMIRHNDRESYNLECGDYI	341	
Db	241	LRALKKNRVLVLEPNPAFVLPETHADAILLVIGGRALMIRHNDRESYNLECGDYI	300	
OY	342	RIPAGTFEYLINRDNNERHLIAKFLQITISTPGQYKEFFPAGQONEPYLSTFSKEILEA	401	
Db	301	RIPAGTFEYLINRDNNERHLIAKFLQITISTPGQYKEFFPAGQONEPYLSTFSKEILEA	360	
OY	402	LNTQAEERLQVIGQOREGYIVSASQOIRELTRDSESRMHITRGGSSRCQPNLFNKR	461	
Db	361	LNTQAEERLQVIGQOREGYIVSASQOIRELTRDSESRMHITRGGSSRCQPNLFNKR	420	
OY	462	PLYSKRYCOAEVEKREDYKROLQDMQVAFIANITDGSMMGPFENFRSPKVVVVAASGADY	521	
Db	421	PLYSKRYCOAEVEKREDYKROLQDMQVAFIANITDGSMMGPFENFRSPKVVVVAASGADY	480	
OY	522	EMACPHLSGRHGGRRGGRKHEEEDVHYEYKARLSKREALVVYVGHVPVYVSSGNENLL	581	
Db	481	EMACPHLSGRHGGRRGGRKHEEEDVHYEYKARLSKREALVVYVGHVPVYVSSGNENLL	540	
OY	582	LEAFCTINONNHENLAGERNVLDQOIEPOAMELFAAPRKVVEELFNSQDSITFFGPR	641	
Db	541	LEAFCTINONNHENLAGERNVLDQOIEPOAMELFAAPRKVVEELFNSQDSITFFGPR	600	
OY	642	OHQOOSSRSTKQOQPLVLSILDFVGF	666	
Db	601	OHQOOSSRSTKQOQPLVLSILDFVGF	625	
RESULT	4			
OYSEW4	ID	PRELIMINARY:	PRT:	593 AA.
AC	OYSEW4			
RC	01-MAY-2000 (TREMBLrel. 13, Created)			
RA	Teuber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.;			
RT	"Identification and cloning of a cDNA encoding a vicilin-like protein,			
RL	Jug r 2, from English walnut kernel (Juglans regia): a major food			
RR	allergen."			
OR	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
OR	EMBL: AF066055; AAF18269.1; --			
OR	HSSP: P02853.2PML			
OR	INTERPRO: IPR001113; --			

PFAM: PF00546: Seedstore\_7s; 1.

DR NON\_TER 1

FT SEQUENCE 593 AA; 69990 MW; 9BA127E19B18C0D8 CRC64;

Query Match 37.5%; Score 1324.5; DR 10; Length 593;

Best Local Similarity 44.3%; Pred. No. 3.6e-88;

Matches 274; Conservative 118; Mismatches 169; Indels 57; Gaps 16;

QY 73 DNQDDPDTCCOQOCHRCQHQESGSPROQOYCCORCKEICEE-----EEEYNRQ-----119

Db 9 ENRPDRFQRYOQRCQCRGQOGQOQOCCQIRCEBRLFEEDQSRQSEERERRRRGRDVDDQN 68

QY 120 -RDPQOQYEQOQCRQQRHETEPRHMQTCQOQCECRERYEKKQKQKRIEEOQREDEEYEE 178

Db 69 PDRPQRYEQOQOQOQER--QRRGOETOTLCRRCEQRHQOEEERERGRGRQO-----117

QY 179 RMKEEDNRDRDQOQREYEDCRRRC--EQOEPQOYQOYOCQ--REOROHGR--GGDLINPQ 234

Db 118 -----DQO-QYHHQRCRQOQOESPERQOCCQRCRQOYKQEQOGRRGGEASPR 167

QY 235 GSGGATYEEGEEKQSDNPYIFEDERSLSTRFRTEBGHISYLENFGRSKILRALKNYRLVL 294

Db 168 ESRGR-----EEEOQRNPYIFFSQISRSRHESEEGEVKLEFTEFTELLRIENYRVIL 224

QY 295 EANPAEFLPRLADATLLVTGGGALKMHRPNRESYNLECGDVI RIPAGTFEYLNR 354

Db 225 DANPNTSMPLPHKKADESAVVTGRATITLVSQETRESFNLECGDVI RVPAGATVYVINO 284

QY 355 DNNERLTIATKFLQTTSPGOYKEFFPAGQNP--EPYLSFESKEITLALNQAERLGVL 413

Db 285 DSNERLEWVKLLQPVNNGQGFREYYAAGAKSPQDSYLVRFSDNLVALNTPRDLERFF 344

QY 414 GQ--QREGVITISAQEOJRELTRD--DSESRKMHIRGGESSRQPYLNFKNRPLYSKYG 469

Db 345 DQOQRESEVILIRASQCKLRALSQHMASGQRPW---GRSSGGPILKSSPSYSQFG 400

QY 470 QAYEYKPEPDYRQOLDMDQSVFANIITQGSNMGPFFNTRSTVYVVVVASGEADVEMACPILS 529

Db 401 QFFEACPEPEHQLEDMQDLVYAEIKRGAMVPHYKNGATVYVYVVGSTGYEMACHYVS 460

QY 530 GRHGRRGGRGKHEEEDY-HTEQYKARLSKEALIVVPGHVPVYVSSGNENLLFAFGIN 588

Db 461 SOSYEGQRRQOEESEESTGRQKVTARLARQDIVIAGHPAIATASQNEMLRLTGPDIN 520

QY 589 AONNHENFLACRRNVLTQOIPQAMELAFAPRKEVELEFSSQDESTFPFGPROHQSS 648

Db 521 GENNQDRDLAG-QNNIINQLERAKELSFNNPREIEIEFESQHESTFVPTER---QSR 575

QY 649 RSTKQOQPLVSLDFVGF 666

Db 576 RGQGRDHPILASILDFAFF 593

RESULT 5

Q9ZWM13 PRELIMINARY; PRT; 810 AA.

AC Q9ZWM13

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)

DE PV100.

OS *Cucurbita maxima* (Pumpkin) (Winter squash).

OC Eukaryota, Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Cucurbitales; Cucurbitaceae; Cucurbita.

OX NCBI\_TaxID=3661;

EN [1]

RP SEQUENCE FROM N. A.

RC STAIN-KUROKAWA AMAKURI NANKIN; TISSUE=COFFLEDON;

XX MEDLINE=99107919; PubMed=9891029;

RT Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.;

"Multiple functional proteins are produced by cleaving Asn-Gln bonds

RT of a single precursor by vacuolar processing enzyme."  
RL J. Biol. Chem. 274:2563-2570(1999).  
DR EMBL: AB019195; BAA34056.1; -  
DR HSSP: P02853; 2PHL.  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
DR PRODOM: PD081059; -; 1.  
SQ SEQUENCE 810 AA; 97314 MW; A829A3F7542266AB CRC64;

Query Match 33.7% Score 1191.5; DB 10; Length 810;  
Best Local Similarity 34.3% Pred No. 2.7e-78;  
Matches 284; Conservative 144; Mismatches 218; Indels 183; Gaps 23;

```
QY 1 MAINTSNLCSLFLSLFLSTVSLAE-----SEF- 31
DB 1 MALSKYKLRCLLAFLLFLLACLSVGLGDKGESLSGAGVDHDCVNRCELKKANNDEFA 60
QY 32 -----DROEYECRQCMQLETS-GQMRVCYSQCKREE----- 65
DB 61 ACKKGGVNOGRSPRAEYECRLRCOVAERGVQQRKCEVCEERLEREREGEDVDEV 120
QY 66 ---DIDMSKY-----DNDPOTDQOQCRRCROE 93
DB 121 ERDPEWEEBQRRREHERERERERERERERERERERERERERERERERERERER 178
QY 94 SGFRQOQYQCRCKELCEEEEEYNRQDPQOQYEQOCORQHETEPRHMQTCQRCERR 153
DB 179 RRRREQERERERERERERERERERERERERERERERERERERERERERERER 232
QY 154 Y-----EKE-----KKQCKRYEEQOREDEBEKEEKKEDDKRD-----QOREYE 195
DB 233 RRGGRDEDEENQDPDMREQOKRKEQERERERERERERERERERERERERERER 292
QY 196 DCRRCRCEOE-----PROOYOCORCRQOROHGSG-- 227
DB 293 ERRRRERERERERERERERERERERERERERERERERERERERERERERER 352
QY 228 -----DLINPORGSGRYEE-----GEKQSDNPYFDEKSLSTRFTEEGHISVL 273
DB 353 BQSRREDERRERERERERERERERERERERERERERERERERERERERERER 412
QY 274 ENFYGSKLRALKNRVLLENPAEVLPTHLDAALLVTGGGALKMHRDNR--- 330
DB 413 ERRSESELKGIKNORLALLEARPHFTFVPHLDACVLLVVRGATITTVQEKRETR 472
QY 331 -SSYNLECGDVIRIPAGTFEYLINRDNNERLHIAKFLQITSPGOYKEFPFAGGONPEPY 389
DB 473 KESYNVESGDVMTIPAGTLLYLANOE-NEDLOIVKLVQYVNNPGEKDYLSAGES-QAY 530
QY 390 LSTFSKEILLEALNTOAERLQVLAQORE--GVIISASQEOIRELTRDSESRMHIRRG 447
DB 531 YSFFSNDVLEALNIPROKLEIFKORRERGGKIYRASQOELRALSO-----RATSVRKG 585
QY 448 GSSSRPVLNFKRPLYSKXGOAYEVKPEDYROLQDMVSVFIANIITGQSMGPFNFR 507
DB 586 SGGVRAPIKLESQTPYNNQYQMEACPEDEFPQLKRTVAISVDIKQGMVNFENSR 645
QY 508 STRVVVVAAGVADENACPHLS-----GRHGRGRGKRHEEE-----DVHYEYVARLS 557
DB 646 AHWVVFVSGASGFEMACHIOSOMQORGRREERHWRREERERERERERERERER 705
QY 558 KREAIIVPVGHVAVVSSGNENLLFAFGINANNHNFELARERNVLOQIEPOAMELAF 617
DB 706 ECGVLVIPAGHIAIMASPNENRLVGFGINANNHNRPLAGRE-NIMNDELBEAKELAF 764
QY 618 AAPKREVEELFNSQDESITFPGRPHQOOSRSTKQOQPLVSLIDFVGF 666
DB 765 NVEGQADEIFRSORESEFTEGPEGRRRST-----ERSPLLSILKLGY 809
```

ID 043358 PRELIMINARY; RRT; 525 AA.  
AC 043358;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE VICILIN PRECURSOR.  
GN CSV.  
OS Theobroma cacao (Cacao).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Malvales; Malvaceae; Theobroma.  
OX NCBI\_TaxID=3641;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAVES;  
RX MEDLINE=92288309; PubMed=1600151;  
RA McHenry L., Filtz P.J.;  
RT "Comparison of the structure and nucleotide sequences of vicilin genes  
of cocoa and cotton raise questions about vicilin evolution.";  
RL Plant Mol. Biol. 18:1173-1176(1992).  
DR EMBL: X62625; CAA44493.1; -  
DR HSSP: P02853; 2PHL.  
DR MENDEL: 30919; Thecc; 1188; 30919.  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
DR PRODOM: PD081059; -; 1.  
KW signal.  
FT SIGNAL.  
FT CHAIN 1 24 POTENTIAL.  
SQ SEQUENCE 525 AA; 60798 MW; 19114CD5C248905D CRC64;

Query Match 29.3% Score 1036; DB 10; Length 525;  
Best Local Similarity 41.4% Pred No. 2.9e-67;  
Matches 213; Conservative 91; Mismatches 156; Indels 54; Gaps 11;

```
QY 109 ICEEEENR---QRDPQOYEQOCORQHETEPRHMQTCQRCRRYKERRKQOKRY 165
DB 22 LCGSAGYGRKQYENDPRQOYEQOCRCSEATEREQECQRCR-----EY 70
QY 166 EQQREDEKYEERKMEENKRDPOOREYEDCRRCRCEOE--PROOYOCORCRQOROH 223
DB 71 KEOQROQEBEL-----QROYQOCQRCQEOQOQOQOQOQOQOQOQOQOQOQOQ 116
QY 224 GRGGLINPORGSGRYEEGEGEKOQSDNPYFDE-RSLSTRFTEEGHISVLENYGRSKL 282
DB 117 ER-GEHENYHNKKNRSEEEGQORNNPYFPKRRSFOTRPFDEBGNFKILORFAENSP 175
QY 283 LRLKNYRLVLEANPAEVLPTHLDAALLVTGGGALKMHRDNRSYNLECGDVIR 342
DB 176 LKGINDYRLAMEANENFTILPHICDAEAIYVTNGKGTITFVTHENKESYVQRGTVS 235
QY 343 IPAGTFEYLINRDNNERLHIAKFLQITSPGOYKEFPFAGGONPEPYLSTFSKEILLEAL 402
DB 236 VPAGSTVYVSDNOEKLITAVLALPVNSPGKYLEFPFAGNKNPEYVGAFTYEVLTVE 295
QY 403 NTOAERLQVLAQORE-----EGVIISASQEOIRELTRDSESRMHIRRGESSRGP 454
DB 296 NTQREKLEELLEQROKROQOQGMFKAKPEQIRAIQOQATSPR-----HNGGE--RLA 349
QY 455 YNLFNKRLPYSNKXGOAYEVKPEDYROLQDMVSVFIANIITGQSMGPFNFRSTRKVVV 514
DB 350 INLLSQPYVSNQNGFEFACPEDEFSQFQNDVAVASAFRLNGALFVPHYNSKATFVVV 409
QY 515 ASGEADVEMACHLSGRHGRGK--RHEEEDV-----HYQVKARLSKREAIIVPV 566
DB 410 TQGYAGQACHPHLSROSQSGOSGRODRREDEESEETFFGFOOVKAPLSGDVFAVA 469
QY 567 GHPVVFVSSGNENLLFAFGINANNHNFELAGR 600
DB 470 GHAVTFPASKDQPLNAVAFGLNANNORITFLAGR 503
```

RESULT  
043358

6

```

RESULT 7
ID 003865 PRELIMINARY; PRT; 582 AA.
AC 003865;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE VICILIN-LIKE EMBRYO STORAGE PROTEIN.
GN GBL1-L.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
OX NCBI_TaxId=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M6A;
RX MEDLINE=92090707; PubMed=1752424;
RA Belanger F.C., Kriz A.L.;
RT "Molecular basis for allelic polymorphism of the maize Globulin-1
  gene.";
RL Genetics 129:863-872(1991).
EMBL: X59083; CAA41809.1; -.
DR HSSP: P50477; ICAU.
DR MENDEL: 11234; Zeama:1188;11234.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 1.
DR PROSITE: PS00867; CPsASE_2; UNKNOWN_1.
DR PRODOM: PD081059; -. 1.
KW Seed storage protein.
SQ SEQUENCE 582 AA; 66162 MW; 680D85FECDCB885 CRC64;

```

Query Match 25.3%; Score 895; DB 10; Length 582;

Best Local Similarity 39.6%; Pred. No. 5.8e-57;

Matches 201; Conservative 83; Mismatches 183; Indels 40; Gaps 11;

```

QY 182 EEDNRKPPQOREYEDCRRCRCEQOEPRQOYOCQRRCRE-----QORHNGSGDLINPORG 235
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 25 EDDNNHNGGKSGGQCYRCEDEPRWNRPRCLRGCRREERKQVSRHNRAD-----RS 79
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 236 GSG-----RYEEGEEKQSD-NPYFDEKSLSTRTERGHSILENTYGNKSKLLRALKNR 230
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 80 GCESSSEDERQEKQKORRPRVYFDRSRFRVYVRSQGLRVLPRPEDEVSRLLGLRDIYR 139
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 291 LVLEENPNAFVLPTHLDADAILLVTGGGALKMIHNDNESYNLGCQVIRIPACTTGY 350
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 140 VAVLEAPRSEFVVPSTDAHCICVAGEGVVTTENGERRSTYIKOGHVFAPAGAVTY 199
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 351 LINDNNERLHIAKFLQTTISTPGQYKFFPAGQNPERYLSTFSKELLEALNTQAERLR 410
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 200 LAMTDGKKLVITIKLHTISVPEGEQFFPGGRNPESFSSKSLQRAAYTSSRLR 259
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 411 GVUGQ--QREGVITISAOEQLRELTRDSE---SRMWHIRGGESSRGPNLFRKPLYS 465
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 260 RLGRHGOQDGIIIVRAITEQTRRELRRHASEGCHGHPRLPFPGE-SRGPYSLLDQRPSTA 318
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 466 NKYGQAEYKPEDEYRQLOMDVSVFANTIOGSMGPFNTRSTKVVVAASGADVEMAC 525
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 319 NQHGQLEADARSFHDLAEVDVSFANTTAGMSAPLFTFRSFKIAYVNGKGAIEIVC 378
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 526 PHLSGRHG-----GRGGKRHEEEDVH-----YEOVKARLKRREAIIVPCHPVY 571
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 379 PRRHOGSGSEERRRDKGRSEEEBESSSEQEBAGOCYHTIRARLSPTGAFVVPACHPEY 438
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 572 FVSSGNENLLFAFGINQNNHNFAGREENVLQOIEPAMELADAPRKEVEELFNSQ 631
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 439 AVASRDSNLIQIVCFEYHADNREKVLGAD-NVLOKLDVAKALSPASKAEDEVILGSR 497
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 632 DESIFPGPRQ---HQOQSSRTKQOO 655
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 498 REKGFLPGPEESGHERREOEEEREE 524
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 8
ID 003678 PRELIMINARY; PRT; 637 AA.
AC 003678;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE FETAL STORAGE PROTEIN.
GN BEG1 OR GBL1.
OS Hordeum vulgare (Barley), and Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
OX NCBI_TaxId=4513, 4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M6A;
RX MEDLINE=93287988; PubMed=8510647;
RA Heck G.R., Chamberlain A.C., Ho T.H.D.;
RT "Barley embryo globulin 1 gene, Beg1: Characterization of cDNA,
  chromosome mapping and regulation of expression.";
RL Mol. Gen. Genet. 239:209-218(1993).
EMBL: M64372; AAA32936.1; -.
DR EMBL: M81719; AAA34269.1; -.
DR HSSP: P02853; 2PHL.
DR MENDEL: 8553; Horvu:1188;8553.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 1.
DR PROSITE: PS00867; CPsASE_2; UNKNOWN_1.
DR PRODOM: PD081059; -. 1.
KW Seed storage protein.
SQ SEQUENCE 637 AA; 72252 MW; F323F4FF99947C3C CRC64;

```

Query Match 24.8%; Score 875; DB 10; Length 637;

Best Local Similarity 35.6%; Pred. No. 1.9e-55;

Matches 227; Conservative 99; Mismatches 206; Indels 106; Gaps 22;

```

QY 70 SKYDNDDDP--TDCCOQCRRCRQESGPRQOYOCQRCKEICEEEEFNRQDDPQOYE 127
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 27 ASHDEDDRGGHSLDQCYQRCQER--PR--YSARCYQEC-----RDDQOQH- 71
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 128 QCEQCRQHEPTEPRHMQTCQRCERREYKQKQRYEEQREDEE---KYEERME 182
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 72 -----GRHDEEQRGRKMGHGEEREHGGRHNGEGEBEHGCRGHNHGERE 125
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 183 EDNRKPPQOREYEDCRRCRCEQOEPRQOYOCQRRCQEQROHNGR---GDLINPORGSG 238
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 126 EERGRG-----HGRHGEDE-----REERGRGGRHNGEGEREDEEGRGRG 165
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 239 RYEEGE--EKQSDN--PYFDEKSLSTRTERGHSILENTPGGRKLLRALKNYLVLL 294
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 166 RREGGERDEEQDSRRPYVFPFRSFRRIQSDHGFYRALRPFQVSRLLGIRIDYVAIIM 225
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 295 EAPNPAFVLPTHLDADAILLVTGGGALKMIHNDNESYNLECGDYIRIPACTTGYLIR 354
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 226 EVNPRAFVVPSTDAHCICVAGEGVVTTENGERRSTYIKOGHVFAPAGAVTY 199
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 355 DNNERLHIAKFLQTTISTPGQYKFFPAGQNPERYLSTFSKELLEALNTQAERLRVLG 414
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 286 DGRRLVIAKILHTISVPGKF-QEL-----SVKFLASLSKRVLAFAFTSDERLERLN 339
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 415 QO-----RQGVITISAOEQLRELTRDSE---SRMWHIRGGESSRGPNLFRKPLYS 465
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 340 QROGQEKTSVSIIVRASQEDQLRELRELAEGGQGHMPPLPFRFGSDRPFNLLQGRPKTA 399
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 466 NKYGQAEYKPEDEYRQLOMDVSVFANTIOGSMGPFNTRSTKVVVAASGADVEMAC 525
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 400 NRGRLYLEADARSFHDLAEVDVSFANTTAGMSAPLFTFRSFKIAYVNGKGAIEIVC 459
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 526 PHL-----SGRHGRRGGRKHEEED-----VHYEYVAKRLSKRE 560
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db 460 PHLGSRSEERENHG--GRRREEDDQOQRRGSESESEEEEDQKRYTVNARSRS 517  
 Qy 561 AIVPVGHPVVEVSS--GNENLLFAFGINAONHENTFLAGERNYLQOIEPOMELAEA 618  
 Db 518 AIVPVGHPVVEVSS--GNENLLFAFGINAONHENTFLAGERNYLQOIEPOMELAEA 618  
 Qy 619 APRKVEEELFNSODEST-FFPGPROHOOSSSSTKOO 655  
 Db 577 RPAREVEYFRAODODEGFVAGP---EQOSRPEOEQO 610  
 RESULT 9  
 ID 09M3X6 PRELIMINARY; PRT: 613 AA.  
 AC 09M3X6;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CONVICTILIN PRECURSOR.  
 GN CVC.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;  
 OC Fabales; Fabaceae; Papilionoideae; Pisum.  
 OX NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BIRTE; TISSUE=SEED COTYLEDON;  
 RA Casey R.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ276875; CAB8285.1; -.  
 KW Signal.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 613 CONVICTILIN.  
 SQ SEQUENCE 613 AA; 72063 MW; 49AED99F6135DD19 CRC64;  
 Query Match 24.1%; Score 851; DB 10; Length 613;  
 Best Local Similarity 30.9%; Pred. No. 9,7e-54;  
 Matches 212; Conservative 135; Mismatches 236; Indels 104; Gaps 14  
 Qy 1 MAINSNLCSLFLSLFLSTTVSLASEFPROYEECKRCMOLETSGMRCVSCQD 60  
 Db 1 MATTKSRPPLLLGLIFLASVCVTYANYD-----EGSEPRVPAQERGRQEE 51  
 Qy 61 KRFEEIDW-SKYNDQDPQDQCOCRCRQESGPRQOYCORH---CKEICEEEE 115  
 Db 52 KEKHNHGRMSYEEEDB---EGQNRGRQEGKEKRRHGRMPSTEKODEBEKOK 107  
 Qy 116 YNRQDPQOQYEQOCERCORHETEPRHMQTCQOCERRYEKERKQOKRYEQOQDEEK 175  
 Db 108 YRYQREKDEBEKQYQYQREKKEQEVQPRERWERREDEQVDEWKGSGRRDPEER 167  
 Qy 176 YEEKKEEDNKRDPQ-QREYEDCRRRCQOQPRQOYCCORRCREQORHGRGDLINPOR 234  
 Db 168 ARLRRREERTKDRRHQRGEEESSSEQERR----- 200  
 Qy 235 GSGRYEGEGEKOQNPYFDERSLSTRPRPEGHISVLENYGSKILLRALKNYRLV 294  
 Db 201 -----NPELFKSNKLLTFPENNGHRLQDFDKSDSLFENLQNRVLEY 245  
 Qy 295 EANPAFVLPFLDADALLVYGSGALKMIRDRRESYNLECGDVIRIPAGTFEYLNR 354  
 Db 246 RAKPHITFLPQHIDADLLVYLSGKAILTVLSPNDRNSYNLERGDTIKLPATTSYLVNQ 305  
 Qy 335 DNNEFLHLAKFLQTIYSTGQYKEFPFAGQNDPEYLSFESKILEALNTQAERLRGVLG 414  
 Db 306 DDEEDRLVLDVLPVNGGKFEAFPLA--KKNQYLRGSKNILLASVNTRETIEKVL 363  
 Qy 415 QQRE-----GVIIASQOIRIELTRDDESESRMHIRRGESSGRYNLFNKR 462  
 Db 364 EOEKDRKROOGEETDIAVKVSRQIEELKLAQSS---KSLPSEFEINIRSHRP 419

[illegible]

ID	AC	Q01674	PRELIMINARY;	PRT;	545 AA.
QY	448	GESSGNPNTLFNFKRPLYSNKKYGOALVEVPEEDYRLOQMDVSVFLANTIGGSMMPFFNTR	507		
Db	339	--SEDKPFLNRKRDPIYSNKKLGKFFETIPKRNPLRDLIDFLSTVDNMEGALLPLPHNSK	397		
QY	508	STRVVVVASGEADVMAACPHTLSGRHGC--RRGCKRHEEEDVHYHYEQVKARLSREALIYVP	565		
Db	398	ALIVILVINEGDAINIELV-----GIKEEQDQEQDQEQEQLVLEKRYKRLSESDIDIVP	448		
QY	566	VGHPIVYVSSCGNENILLEFAGIMANONHENFLAGERNVYLQOITEPAMELFAAPRKVE	625		
Db	449	AGYPVVV--NMTSNLNFPAIGAINENNRNPLAGSODNVISQIPSOVQELAFPGSAQAVE	506		
QY	626	ELFNSSODESIFPPGPGROHGOOSNRSTKQOPLVSLI	661		
Db	507	KLLKNQRESYFVDAPKKKEGKGRK--GPLSTIL	540		
RESULT	11	Q01674			
ID	AC	Q01674	PRELIMINARY;	PRT;	545 AA.
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)			
DE	CONVICITILIN PRECURSOR.				
OS	Vicia narponensis.				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Manuoliophyta; eudicotyledons: core eudicots; Rosidae; eurosoids I;				
OC	Fabales; Fabaceae; Papilionoideae; Vicia.				
OX	NCBI_TaxID=3912;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=COVYLEDON;				
RA	Koch G., Koenig S., Becker C., Horstmann C., Schlesier B.;				
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; 2711986; CAA96513.1; -.				
DR	HSSP; P50477; ICAU.				
DR	MENDEL; 12432; Vicia;1188;12432.				
DR	INTERPRO; IPR001113; -.				
DR	PFAM; PF00546; Seedstore_7s; 1.				
DR	PRODOM; PD081059; -; 1.				
KW	Signal; Seed storage protein.				
FT	SIGNAL	1 28	POTENTIAL.		
FT	CHAIN	29 545	75 GLOBULIN, CONVICITILIN.		
SO	SEQUENCE	545 AA; 62810 MW; 459A876F925E5A87E CRC64;			

	Query Match	Similarity	23.2%	Score 818	DB 10	Length 545
Best Local	Similarity	34.3%	Pred. No. 2.1e-51			
Matches 196	Conservative	101	Mismatches 204	Indels 70	Gaps 14	
QY	109	ICEEEEEYNRRDPPOOOYECOCERCORHETEPERHNOOTCOOCERERYEKEKKRQOORVEE	167			
DB	23	LCVTYANVDEGTETPVPGGRERGRQGEKEEKRH-----GEMRSHHEKAOQGRREWM	75			
QY	168	QOREBEKYEEMKEDNKRDPQOREYEDCRRCEQOEPRQOYOCORCRCEQOROHGRGG	227			
DB	76	ETSEEERADDEMRKQSORHEDPEERARE--RRAEERERRRQME-----	117			
QY	228	DLINFGGSGRGYEEGEEKROSD--NPYYEDERSLSTRFTEGHSIVLENFYGRSKILRA	285			
DB	118	-----GEEKEGSSKSQERRRNPFLFSKMKFTLTFENENGHILRLQRFDRKSLFLFN	167			
QY	286	LKNYPLVLLLEANPNAFVLEPHLIDADAILLVITGGALKNIHNDNRESYVLECGDVIRIPA	345			
DB	168	LQNYALVEYRARAPHILFLFOHIDADLILVLBSGRAILTVLSPNDRSNYSVLEGGDTIKLPA	227			
QY	346	GTFEFLINRDNNEHTHIAMKFLQTSIPGQYKEFPAGGQGNPEPYLSTSEKELTEALNQ	405			
DB	228	GTTTLLNODDEDELRVYDLSTISVNPCKVESEFGISGSKN--QYLRGSKNLLSEASLVNK	285			
QY	406	AERLGVL-----GOORE-----GVIIASQEOJELFELTRDSESRWRHIGGEES	451			

```

Db      266 YETIEKVLLEEPQSIGCKRRSQOETNALVYKSREDOVEELKRLAKSS-----KKGVSE 341
OY      452 RGPYLLFNKRPLYSNKYQOAYEVAPR-DYRQLODMDVSVFIANTITGQSMGPFENRSTK 510
Db      342 FEPFLRQONKYSKRSFKLLEIFPEKKYDLOALDDIFVSSVEINFGMLPHYSRAIV 401
OY      511 VVVVASGADYEMACPHLSGRHGRGCKRIIEEDVHYEDVKRLSKRELIVYVGPV 570
Db      402 ILLVNEGGLNLE-----LVGAKNEQOEDEREDEBOV--QRYEARLSPGDVIIIPACHV 454
OY      571 VVSSGNNLLLEAFGLIAQNNHNHETLAGRENNYLOOIEPAMELAEAPRKVEEELNS 630
Db      455 AVSASSNNLLL--GGINAENQNNRNLFTGSDNVITQIENPVKELTFPGSAQEVNRLKN 512
OY      631 ODESTFFPGPHQHQOQSSRSKTKQOQPLVSTL 661
Db      513 QEHSHF--ANAEPQCKGEESQKRSKPSSTIL 541

```

RESULT	12			
Q9SP11				
ID	Q9SP11	PRELIMINARY;	PRT;	489 AA.
AC	Q9SP11;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)			
DE	SUCROSE BINDING PROTEIN HOMOLOG S-64.			
GN	SBP.			
OS	Glycine max (Soybean).			
OC	Eurycotylae: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:			
OC	Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I;			
OC	Fabales; Fabaceae; Papilionoideae; Glycine.			
OX	NCBI_TaxID=3847;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Pedra J.H.F., Delu-Filho N., Pirovani C.P., Contim L.S., Dewey R.E.,			
RA	Otoni W.C., Fontes E.P.B.;			
RT	"Antisense and sense expression of a sucrose binding protein homologue			
RT	gene from soybean in transgenic tobacco affects plant growth and			
RT	carbohydrate partitioning in leaves."			
RL	Plant Sci. 0:0-0(1999).			
DR	EMBL: AF191299; AAF05723.1; -			
DR	HSSP: P02853; 2PHL			
DR	INTERPRO: IPR001113; -			
DR	PFAM: PF00546; Seedstore_7s; 1.			
SO	SEQUENCE 489 AA: 55834 MW; 99BC0D45EDECCD2 CRC64;			

[illegible]



Db 307 TFSNGGRLETEGVPDEKSMQLRLMLTFNTQSRMSTIHVNSHATKALVMDRGHL 366  
522 EMACPHLSGHRGGRGKRHEEEDVHYEQVKARLSKREAIYVPGHVPVSSGNENLL 591  
Db 367 QISCPHMSRSBSK-----HDKSSPSYHRISADLPGAVFVPPCHPPTTASKNENLL 420  
582 LFAFGINAONNHENFLAGERNVLOQIEPOMELAFAPRKEVELLFNSODESIFPPGPR 641  
Db 421 IICEFVNVADNKKFTFAGKD-NIVSSLDNVAKELAFNPSEMVNGVF-ERKESLFFPEPFL 478  
Qy 642 QHQOOSRS 650  
Db 479 PSEERGRA 487

RESULT 13  
ID 049927 PRELIMINARY: PRT: 483 AA.  
AC 049927;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE P54 PROTEIN.  
GN P54.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Castillo J., Marquez J.A., Franco L., Ballester E., Rodrigo M.I.;  
RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Y11207; CAA72090.1; -  
DR HSSP: P50477; ICAU.  
DR MENDEL: 28319; Pissa: 3005; 28319.  
DR INTERPRO: IPR001005; -  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
DR PROSITE: PS00041; HTL\_ARAC\_FAMILY\_1; UNKNOWN\_1.  
DR PRODOM: PD081059; -; 1.  
SQ SEQUENCE 483 AA; 54662 MW; 8127BDAAD178F3D CRC64;

Query Match 21.8%; Score 770.5; DB 10; Length 483;  
Best Local Similarity 31.6%; Pred. No. 4.9e-48;  
Matches 163; Conservative 113; Mismatches 169; Indels 71; Gaps 12;

Qy 135 RHETPRHMQTQOQRCERYEKERKQOQRYEQOQREDEKTEENKEDKRDPOQREY 194  
Db 28 RREKDP-ELTTCKDQCD-----MORQYDEDK----- 53  
Qy 195 EDCRRRCQEQEPRQOYQOQRCQOQROHGRGDDLINPQRCSGRYEEGEKOSDPYVF 254  
Db 54 RICEKCD-----DIKKQERQKHKEH-----EEEEQEQDEDEDPYVF 93  
Qy 255 DERLSSTRETEGHSIVLENFYGRSKLLRALKNRYLLLEANPNAFVPTHLADAILL 314  
Db 94 EDNDEFTLDTDGRVLLINKENESKLLKNIENTGLAVLETKANAFSLPHYDSAILLF 153  
Qy 315 VMGRCALKMIHRDNRESYNLECGDVIRIPAGTFFVLIHRDNNEHLIAKF--LQITST 371  
Db 154 NIKGRGIIQLVADERTERENLEEGDIMRVPAQTPYLVNRDEKELYIAFIHPPSSGA 213  
Qy 372 PQQYKEFPAGQONPEPYLSTFSKLEALNTQAERLRGVIGQOREGYITASQOQIRE 431  
Db 214 PVNLEFEFFSARKEPSVYLNFTSSKYLQALAKSSKGLFTVLDEQKGRIFKIEKDVNG 273  
Qy 432 LTRDSESRMHIIRGGESSRGPNYLNFKRPLYSNKYGAYEVKPEPDYQ-LQMDVSVF 490  
Db 274 LPPKKS---LWPF---GGFFKSPFNIFSNPAPSNKFGSLFEVGSQEKSGLEGLMLLT 327

Qy 491 IANITQSMKGFENRSTRKVVVAVSGEADVEMACPHLSGRHGRGKRHEEEDVHYE 550  
Db 328 LANITGSMSTIHTNNAKNALVIDEGELEMACPHMSSSSNSR-----QKSSISYH 382  
Qy 551 QYKARLSKREAIYVPGHVPVSSGNENLLFAFINNQNHNENFLAGERNVLOQIEP 610  
Db 383 MNAKLPGVMFVVPAGHFPFNIAASKKNLIYCFEVAQNRKNKALAKG-KNIVSALDK 441  
Qy 611 QAMELFAAPRKEVELLFNSODESIFPPGPRHOQO 646  
Db 442 AAKEVAFDIAAEKVEDEFEKKE-FFFPYDNEKKE 476

RESULT 14  
ID 040873 PRELIMINARY: PRT: 448 AA.  
AC 040873;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE VICILIN-LIKE STORAGE PROTEIN.  
OS Picea glauca (White spruce).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Coniferopsida; Coniferales; Pinaceae; Picea.  
OX NCBI\_TaxID=3330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PG118; TISSUE-SOMATIC EMBRYO;  
RX MEDLINE=93004485; PubMed=1391775;  
RA Newton C.H., Flinn B.S., Sutton B.C.;  
RT "Vicilin-like seed storage proteins in the gymnosperm Interior spruce  
("Picea glauca/engelmannii)".  
RL Plant Mol. Biol. 20:315-322(1992).  
DR EMBL: X63191; CAA44873.1; -  
DR HSSP: P02853; 2PHL.  
DR MENDEL: 12115; PICGL1188; 12115.  
DR INTERPRO: IPR001113; -  
DR PFAM: PF00546; Seedstore\_7s; 1.  
DR PRODOM: PD081059; -; 1.  
KW Seed storage protein.  
SQ SEQUENCE 448 AA; 50200 MW; 467AE5BA12E6D157 CRC64;

Query Match 21.7%; Score 765; DB 10; Length 448;  
Best Local Similarity 37.8%; Pred. No. 1.1e-47;  
Matches 162; Conservative 89; Mismatches 154; Indels 24; Gaps 9;

Qy 232 PQRCSGRYEEGEKOSDPYFDEKSLSTRRTTEGHSIVLENFYGRSKLLRALKNRYL 291  
Db 34 PLYLGRGRGR-EEREENPYVPHSDSFTFRASSSEGETRALPNFGEVSELLEGIRKFRV 92  
Qy 292 VLEAPNPAFVLPPTHLDADAILLVYTGRCALKMIHRDNRESYNLECGDVIRIPAGTFFV 351  
Db 93 TCLEMKPNTVMLPHYTDAILVYTRGRYIAYVHONELYKRLLEGVDFGVPSCHTFL 152  
Qy 352 INKDNNEHLIAKFLQITST-PQOYKEFPAGQONPEPYLSTFSKLEALNTQAERLR 410  
Db 153 VNNDHNTLRISLVRPVSTVGEQPFVAGGRNQTYASAFSDVLEAANTVQOLE 212  
Qy 411 GVLGQOREGYITISASQOQIRELTRDSESRMHIIRG--GESSR---GPNYLNKRP 463  
Db 213 RIFGGRKSGVITHANDEQIREMR-----KRGSAGSMSAPENHPENLNQKPD 262  
Qy 464 YSNKGOAEEVAPEDYRQODMDVSYFANINQOSMGMPFTTRSTKYVYVAVSGEDVEM 523  
Db 263 FENNEKRTTIAPKNTPTLDALDVSYGLADLPNGSTJAPSLNKSSTSGITVNGGRLEM 322  
Qy 524 ACPHLSGRHRGGRGKRHEEEDVHYEQVKARLSKREAIYVPGHVPVSSGNENLLLF 583  
Db 323 ACPHL-GQHGW---SSPREGDDDITYQVMAKLRTGSYIYVAGHPITFIATNSRLQIL 379  
Qy 584 AFGINAONNHENFLAGERNVLOQIEPOMELAFAPR-KEYVELLFNSODESIFPPGPR 642



Db	380	MFIDLNTKRNENQFLACK - NNVLNLTLEKEIKQLSTNVFRGEETIEFLVLAQKDQVYLKRPQR	438
QY	643	HQQQSSRST	651
		: : : :	
Db	439	RSRDEARSS	447

Job time: 1557 sec

RESULT	15		
022121			
ID	022121	PRELIMINARY;	PTT: 417 AA.
AC	022121;		
DT	01-JAN-1998 (TEMBurel, 05, Created)		
DT	01-JAN-1998 (TEMBurel, 05, last sequence update)		
DT	01-OCT-2000 (TEMBurel, 15, last annotation update)		
DE	BEA SUBUNIT OF BETA CONEALYCININ.		
OS	Glycine max (Soybean).		
OC	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta		
OC	Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eustosids I;		
OC	Fabales; Fabaceae; Papilionoideae; Glycine.		
OX	NCBI_TaxID=3847;		
RN	[1]		
RF	SEQUENCE FROM N.A.		
RC	STRAIN=CV. WASESUZUNARI;		
RA	Maruyama N., Katsube T., Wada Y., De La Rosa A., Usumi S.;		
RL	Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AB008679; BAA23361.1; -.		
DR	IISSP; P02853; 2PBL.		
DR	MENDEL; 25075; Glyma;1188;25075.		
DR	INTERPRO; IPR001113; -.		
DR	PFAM; PF00546; Seedstore_7s; 1.		
DR	PRODOM; PD081059; -; 1.		
SO	SEQUENCE 417 AA; 48106 MW; DDEE50B5B779556E CRC64;		

Query Match	21.4%	Score 757;	DB 10;	Length 417;
Best Local Similarity	37.6%	Pred. NO.	3.8e-47;	
Matches 160;	Conservative	89;	Mismatches 150;	Indels 26;
				Gaps 7;

OY	246	KOSNPYVF--DERELSTRFRTEEGHISVLEENFYGRSKLTPALKNRYLVLEAPNVAFVP	304
Dd	6	EDENNPFYSSNFOFLFENONCRITILLOFPKRSQOLENLADRYLVQGSRPNTILLP	65
OY	305	THLDADAILLVTGGRALKMTHRDNRESYLCEGDVIPIAGTTFEVLINRDNNERLHIAK	364
Dd	66	HHAADAFLLFEVLSGRALLTVLNDDRDSYNLHGDAORIPAGTYYVLVNPHDQNLIKIK	125
OY	365	FLOIFSPRGYKEEFPPAGGNPREYLTSTESKELLEALNQARLRGLV-----GGORE	418
Dd	126	LAIPIVNPGRIDDFELFSTQAQSYLOGFSHNILETSFHSEFEIIRKVLYFGEEBEOROOE	185
OY	419	GVLISAQEOELIRELTRDSESRRMHIRGCESSRGPNLNRKPPLYSNKYGQAYEVKPED	478
Dd	186	GVILELSEKDRIOLSRANKSSRKTI----SSDEPRNLRSRMPITSNNNGKFETLPPEK	241
OY	479	YROLQMDVSVFLANITOGSMGCPFFENTRSTKVVVVAASGEADVEMACPHLSGHHG--RR	536
Dd	242	NPOLRIDDLIFLSDVINIEGALLPHFNKSAILVILVINEGANIELV-----GIKEQ	292
OY	537	GGRHEEEDVEHYEQVARKSLREAVIVVRPGHVVPVASSGENULLFARGINNQNHHNF	596
Dd	293	QOKOQOEPELEVOIRKAELSDDVFIIPAAY--PVVATNSMLNFLAGFINENNOQRNF	350
OY	597	LAGEERNVLOOIEPOMELFAAPRKEVELLFNSODESIFFPGRPROHQOOSSRSTKOOP	656
Dd	351	LAGEKDVWVBOIEROVQELAFPGSAQOVVERLLKKOBSYRVVDNAQPOQKEGSGSGR--GP	408
OY	657	LVSTL 661	
Dd	409	FPSTL 413	

